Establishing construct validity of a virtual-reality training simulator for hysteroscopy via a multimetric scoring system

Bajka, M ; Tuchschmid, S ; Fink, D ; Székely, G ; Harders, M

Abstract: BACKGROUND: The aims of this study are to determine construct validity for the HystSim virtual-reality (VR) training simulator for hysteroscopy via a new multimetric scoring system (MMSS) and to explore learning curves for both novices and experienced surgeons. METHODS: Fifteen relevant metrics had been identified for diagnostic hysteroscopy by means of hierarchical task decomposition. They were grouped into four modules (visualization, ergonomics, safety, and fluid handling) and individually weighted, building the MMSS for this study. In a first step, 24 novice medical students and 12 experienced gynecologists went through a self-paced teaching tutorial, in which all participants received clearly stated goals and instructions on how to carry out hysteroscopic procedures properly for this study. All subjects performed five repeated trials on two different exercises on HystSim (exploration and diagnosis exercises). After each trial the results were presented to the participants in the form of an automated objective feedback report (AOFR). Construct validity for the MMSS and learning curves were investigated by comparing the performance between novices and experienced surgeons and in between the repeated trials. To study the effect of repeated practice, 23 of the novices returned 2 weeks later for a second training session. RESULTS: Comparing novices with the experienced group, the ergonomics and fluid handling modules resulted in construct validity, while the visualization module did not, and for the safety module the experienced group even scored significantly lower than novices in both exercises. The overall score showed only construct validity when the safety module was excluded. Concerning learning curves, all subjects improved significantly during the training on HystSim, with clear indication that the second training session was beneficial for novice surgeons. CONCLUSIONS: Construct validity for HystSim has been established for different modules of VR metrics on a new MMSS developed for diagnostic hysteroscopy. Careful refinement and further testing of metrics and scores is required before using them as assessment tools for operative skills.

DOI: https://doi.org/10.1007/s00464-009-0582-4

Surg Endosc (2010) 24:79–88
DOI 10.1007/s00464-009-0582-4

Establishing construct validity of a virtual-reality training simulator for hysteroscopy via a multimetric scoring system

Michael Bajka · Stefan Tuchschmid · Daniel Fink · Gábor Székely · Matthias Harders

Received: 5 February 2009 / Accepted: 25 April 2009 / Published online: 24 June 2009

Abstract
Background The aims of this study are to determine construct validity for the HystSim virtual-reality (VR) training simulator for hysteroscopy via a new multimetric scoring system (MMSS) and to explore learning curves for both novices and experienced surgeons.
Methods Fifteen relevant metrics had been identified for diagnostic hysteroscopy by means of hierarchical task decomposition. They were grouped into four modules (visualization, ergonomics, safety, and fluid handling) and individually weighted, building the MMSS for this study. In a first step, 24 novice medical students and 12 experienced gynecologists went through a self-paced teaching tutorial, in which all participants received clearly stated goals and instructions on how to carry out hysteroscopic procedures properly for this study. All subjects performed five repeated trials on two different exercises on HystSim (exploration and diagnosis exercises). After each trial the results were presented to the participants in the form of an automated objective feedback report (AOFR). Construct validity for the MMSS and learning curves were investigated by comparing the performance between novices and experienced surgeons and in between the repeated trials. To study the effect of repeated practice, 23 of the novices returned 2 weeks later for a second training session.
Results Comparing novices with the experienced group, the ergonomics and fluid handling modules resulted in construct validity, while the visualization module did not, and for the safety module the experienced group even scored significantly lower than novices in both exercises. The overall score showed only construct validity when the safety module was excluded. Concerning learning curves, all subjects improved significantly during the training on HystSim, with clear indication that the second training session was beneficial for novice surgeons.
Conclusions Construct validity for HystSim has been established for different modules of VR metrics on a new MMSS developed for diagnostic hysteroscopy. Careful refinement and further testing of metrics and scores is required before using them as assessment tools for operative skills.

Keywords Virtual reality · Training · Simulation · Hysteroscopy · Evaluation · Construct validity · Learning curves

In the last decade, high-fidelity virtual-reality (VR) simulators have emerged as valuable alternatives for practical surgical skills training [1–6], excluding any risk to cause harm to an individual [7]. Past efforts to incorporate simulation into surgical curricula for laparoscopy provide a valuable roadmap on how a simulator for hysteroscopy could be evaluated, validated, and finally integrated into the training curriculum for gynecology [8].

As a proposed first step in the validation cascade [9], face validity has been established with high ratings for both realism and training capacity for HystSim [10], a new surgery simulator for diagnostic and operative hysteroscopy [11]. The presented results demonstrate that potential trainees and trainers accept HystSim as a realistic and useful tool for the training of hysteroscopic interventions.

M. Bajka (✉) · D. Fink
Division of Gynaecology, Department of OB/GYN, University Hospital Zurich, Zurich 8091, Switzerland
e-mail: michael.bajka@hin.ch

S. Tuchschmid · G. Székely · M. Harders
ETH Zurich, Zurich 8092, Switzerland

As a second step of validation, construct validity is usually investigated. Typically, it is established by comparing the performance for groups of surgeons with different degrees of experience [12–19]. The hypothesis is tested that performance scores derived for a certain task on the simulator are significantly higher for experts than for novices.

While it is useful to know whether the different parameters show construct validity, the final goal is to judge and predict performance and ultimately the outcome of an intervention. The first VR simulators in surgery (e.g., the MIST VR [20]) presented abstract tasks with geometric bodies in a synthetic environment, using single criterions such as time to complete a trial or counting of errors, for both validation and assessment. However, it is doubtful whether these common metrics are sufficient to assess surgical performance comprehensively [21–25]. Recent high-fidelity simulators present very realistic simulated surgical scenes, implementing more and more combinations of metrics and scoring systems which express common clinical skills, e.g., "economy of movements" [17] or "precision" [25]. Mackay found that the process of assessing technical abilities is more robust if candidates are tested on multiple parameters using a variety of measures [26]. Van Dongen concluded that the implementation of a scoring system enabled them to assess further aspects of performance [25].

The selected metrics and the superimposed scoring and grading system have to fulfill the following properties: (1) clinical relevance—the metrics have to be as outcome specific as possible, with clear reference to the underlying goal of the procedure; (2) balance—the scoring system should balance well between the sometimes conflicting goals for the metrics, e.g., it should not be possible to compensate low quality of performance with a short intervention time; and (3) simplicity—the feedback should be simple enough to be explained in a few seconds while still providing useful and purposeful guidance for the trainee.

Based on the characteristics above, it becomes clear that each surgical procedure requires a customized scoring system. While some of the metrics apply to most surgeries and are employed by other simulators as well (e.g., intervention time and instrument path lengths), others are unique to hysteroscopy (e.g., time of insufficient expansion of the uterine cavity and visualization of the tubal orifices). Overviews of metrics used by different vendors of laparoscopy simulators can be found in the literature [17, 27]. However, only a few of them have been validated rigorously as assessment tools.

Therefore, analogously to Cao [28], we have performed a hierarchical task decomposition of diagnostic hysteroscopy [29], defining 4 tasks, 15 subtasks, 33 steps, and 46 substeps. This process resulted in the identification of 15 metrics for VR skills assessment which will be used here to develop and validate a multimetric scoring system (MMSS) for diagnostic hysteroscopy.

Thus, the main goal of this study was to explore construct validity of HystSim, i.e., to what extent hysteroscopy simulation in HystSim identifies the quality, ability, and trait it was designed to measure [30]. Since the prime motivation for using simulation is to accelerate the learning of surgical skills, we were also interested in the learning curves of trainees to find out more about the training effect while using HystSim.

Materials and methods

Subjects

The group of novices consisted of 24 medical students with no prior experience in hysteroscopy. They were recruited by an email campaign to medical students in the fifth and sixth years at the University of Zurich, Switzerland. In addition, 12 gynecologists known as experienced hysteroscopic surgeons with many years of practical experience replied to an email invitation to participate in this study.

Apparatus

As in the previous face validity study [10], the HystSim consisted of an adapted hysteroscope (10-mm resectoscope), a virtual patient robot, and the simulation software. The simulation software ran on standard personal computer (PC) hardware (dual 3.0-GHz Pentium processor, 2 GB RAM, NVIDIA 8800 graphics card). The adapted resectoscope tracked all actions and movements of the trainee and was used as input to adapt the simulation accordingly. For this study, only the HystSim diagnostic hysteroscopy software module (version 0.12) was employed. There was no haptic feedback to guide the user in this version. Figure 1 shows screenshots of the running simulation and the hardware setup used in this study.

Multimetric scoring system

Table 1 presents the metrics used in this study, together with a short description of each parameter. Two surgical experts, each having performed more than 500 hysteroscopic interventions, were responsible for defining, weighting, integration, and configuration of the metrics into the scoring and grading system.

For diagnostic hysteroscopy, we grouped the parameters into four modules, i.e., visualization, ergonomics, fluid handling, and safety. The fluid handling module was only

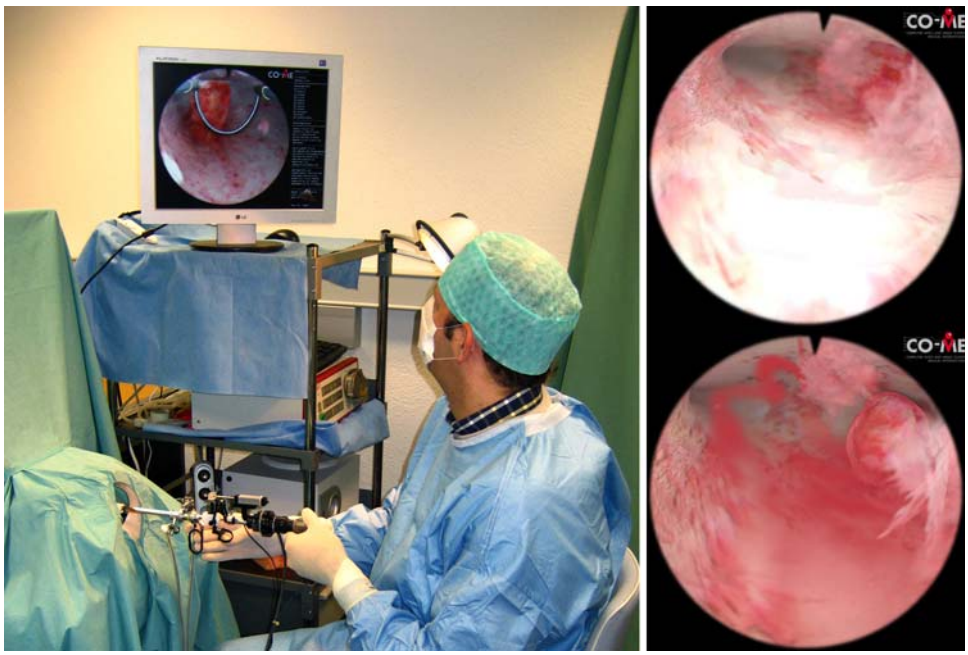

Fig. 1 Hardware setup of the HystSim hysteroscopy simulator used in this study (*left*) and screenshots from the running simulation (*right*)

Table 1 VR metrics for the evaluation of diagnostic hysteroscopy

Parameter	Description
Visualized surface [%]	Percentage of uterine surface which has been clearly visible in the endoscopic view
Left tube visualized [s]	Duration in seconds that the checkpoint in the left tubal orifices has been clearly visible in the endoscopic view. Requires 90° clockwise rotation of the 30° angled scope
Right tube visualized [s]	Duration in seconds that the checkpoint in the right tubal orifices has been clearly visible in the endoscopic view. Requires 90° counterclockwise rotation of the 30° angled scope
Upper cavum visualized [s]	Duration in seconds that the checkpoint at the isthmic part of the anterior wall has been visualized. Requires 180° rotation of the scope
Time out of focus [s]	Duration in seconds that the image focus has been off by more than an expert-defined threshold.
Intervention time [s]	Duration of the intervention in seconds
View horizon unstable [s]	Duration that the horizon defined by the two tubal orifices has been rotated in the endoscopic view by more than 10°, in seconds
Path length [mm]	Distance that the endoscopic camera on the tip of the scope has been moved in millimeters. A short path length indicates proficient tool handling economics
Tool rotation sum [°]	Angular path length (sum of all rotation) in degrees. A low rotation sum indicates proficient tool handling economics
Translation switches [integer]	Number of times that the translation of the instrument has changed its direction. A high number of switches indicates poor handling of the instrument ("sawing style")
Time colliding [s]	Duration in seconds that the endoscopic camera has been in contact with the uterine surface
Time view obscured [s]	Duration in seconds that the endoscopic view has been obscured due to bleeding
Time uterus collapsed [s]	Duration in seconds that the hydrometra of the uterine cavity was not maintained due to rinsing or incorrect settings of the valves. Adequate pressure settings and a proper distension of the uterine cavity indicate proficient fluid handling
Number of spoil cycles [integer]	Number of times that the pressure of the uterine cavity has been changed from low to high pressure. A low number indicates the use of continuous-flow technique, which is appropriate to use while coping with constant bleeding
Distension media needed [ml]	Amount of distension fluid used in milliliters. A low amount indicates proficient fluid handling

employed in the second exercise. For each parameter, the experts set upper and lower limits. Depending on whether the desirable value was low or high, performing above the upper limit resulted in the maximum or a 0 score, while performance below the lower limit resulted in the opposite extremal score. Any value in between was linearly

Table 2 Multimetric scoring system (MMSS) used for both modular and overall performance assessment based on the implemented metrics

Scoring and grading	Max score	Upper value	Lower value	Best
Visualization	70			
Visualized surface [%]	30	85	70	High
Left tube visualized [s]	10	1	0	High
Right tube visualized [s]	10	1	0	High
Upper cavum visualized [s]	10	1	0	High
Time out of focus [s]	10	30	5	Low
Ergonomics	50			
Intervention time [s]	20	180	90	Low
View horizon unstable [s]	10	60	20	Low
Path length [mm]	10	1500	500	Low
Tool rotation sum [deg]	5	3000	500	Low
Translation switches [integer]	5	30	3	Low
Safety	20			
Time colliding [s]	20	12	0	Low
Fluid handling	50			
Time view obscured [s]	25	45	10	Low
Time uterus collapsed [s]	10	20	5	Low
Number of spoil cycles [integer]	5	20	5	Low
Distension media needed [ml]	10	500	200	Low
Overall score	190			
%	100			
Grading	E <60, D 60–69, C 70–79, B 80–89, A 90–100			

interpolated. The maximum score of each metric and therefore its weight was determined by the experts. The resulting scoring chart is shown in Table 2. In order to group the overall scores, grading with letters from A (best) to E (worst) was used.

It is important to note that some values would be meaningful in different groups; e.g. "time uterine cavity collapsed" could be part of both the safety (for measuring risk of movement in the collapsed cavity) and fluid handling modules (for measuring skill to establish the hydrometra). Since the former is already implicitly covered by "time colliding", it was decided to assign this metric to the fluid handling module. It should be mentioned that only one metric ("time colliding") remained in the safety module.

Training curriculum

In order to mimic the actual use of the simulator in a training curriculum and to establish baseline theoretical knowledge, a self-guided, self-paced teaching tutorial was developed. The tutorial included didactic content on uterine anatomy, instruments, and fluid handling as well as hints for navigation inside the cavity and the safe use of the 30° angled scope. The tutorial further provided several movies showing in parallel the endoscopic view and the outside view of the instrument with proposed handling (Fig. 2).

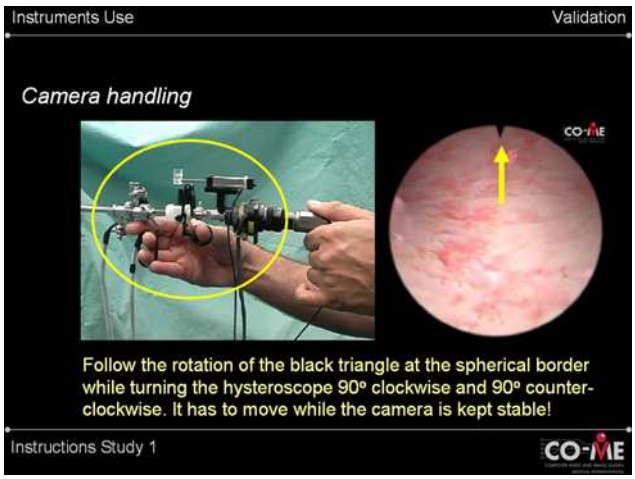

Fig. 2 Screenshot from the self-paced teaching presentation explaining correct camera handling

The goals of the two exercises, and guidelines on how to carry out the task and to avoid complications, were clearly stated before the simulation. An engineer acted as the overseer of the study and started all tutorials and simulations, excluding any further medical knowledge transfer. In the first virtual case—the "exploration exercise"—fluid handling was automatically controlled, the hydrometra was already established, and the endoscopic view was always clear, thus allowing the trainee to fully focus on navigation.

In the second virtual scene—the "diagnosis exercise"—a complete diagnostic hysteroscopy was to be performed, including establishing and maintaining clear view, visualizing the entire cavity with safe use of the angled optics, describing the pathology seen, and reacting adequately to minor complications such as bleeding.

Learning curves

The exploration exercise was carried out five times before moving onto the diagnosis exercise, which was also repeated to a total of five trials. After each trial, the results of the MMSS with goal values for all metrics were displayed in form of an automated objective feedback report (AOFR) (Fig. 3) to further stimulate learning. On demand, further explanations were provided on the scoring system.

In the literature, the first trial on the simulator is often not taken into account because it is believed that its main use is to get the subjects accustomed to the simulation [20, 25, 31–33]. Nevertheless, we believe it is important to register the first contact with the simulator also since it states the absolute starting point of a trainee and leads to the same well-defined study conditions for all participants and all trials. Therefore, in order to quantify the training effect, we compared both the first and the second trial to the fifth trial for the exploration and diagnosis exercises.

All novices were invited to participate in a second training session after a few days break in order to study the effect of repeated practice on the learning curves. Twenty-three out of the 24 novices returned on average 2 weeks later (range 7–18 days). The training curriculum for the second session was identical to the first session.

Statistics

Data were analyzed using the Statistical Package for the Social Sciences (SPSS) version 14.0 (SPSS Inc., Chicago IL, USA) for Windows. To compare novice and experienced scores and in order to compare between two different trials, comparisons were made using the Mann–Whitney *U* test to check for the significance. A *p*-value of less than 0.05 was considered as significant.

Results

Demographics

The median age of the 12 experienced surgeons was 42 years (range 37–69 years), whereas the median age of the 24 novices was 26 years (range 22–34 years). Three of the experienced surgeons and 15 of the novices were female. Nine of the experienced surgeons had previously performed more than 200, two between 101 and 200, and one between 50 and 100 hysteroscopies. None of the novices had performed any hysteroscopies. Seven of the experienced surgeons had mastered more than five complications (heavy bleeding, perforation, fluid overload syndrome), four had previously mastered one to five, and one had not yet mastered any complication. Concerning surgical fitness, five of the experienced surgeons had

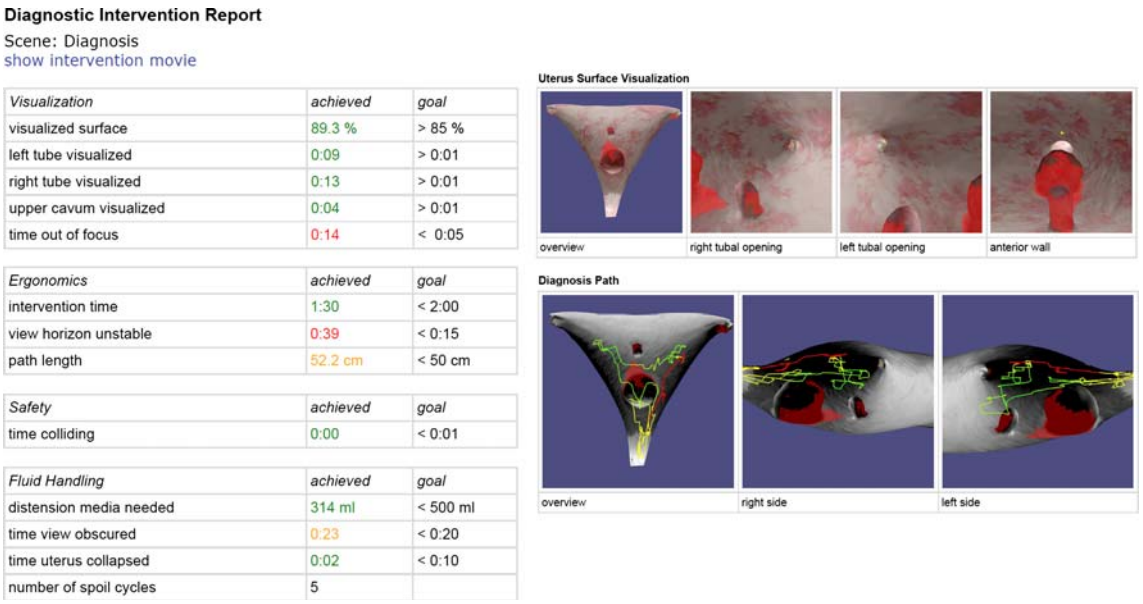

Fig. 3 Automated objective feedback report (AOFR) presented to the trainees after each trial

performed one or more hysteroscopies during the week before the experiments, whereas seven had not.

Construct validity

For each subject, the mean score from trial 1 to trial 5 was calculated for each module of each exercise separately and also for the overall score. The scores from all novices were then compared with the scores of all experienced surgeons using the Mann–Whitney *U* test. The results are shown in Table 3. While the score for the visualization module and the overall score were not significantly different for either exercise, the ergonomics and safety modules for the exploration exercise, and the ergonomics, safety, and fluid handling modules for the diagnosis exercise resulted in highly significant differences. For the ergonomics and fluid handling modules, the experienced group scored significantly higher, but for the safety module novices scored significantly higher in both exercises.

Table 3 Differences between novices (NOV) and the experienced group (EXP) for all module scores for both the exploration exercise and diagnosis exercise

Module	Comparison groups	*P*-value	Result
Exploration exercise			
Visualization	NOV <> EXP	0.179	n.s.
Ergonomics	NOV <> EXP	0.001*	EXP higher
Safety	NOV <> EXP	0.002*	NOV higher
Overall	NOV <> EXP	0.441	n.s.
Diagnosis exercise			
Visualization	NOV <> EXP	0.283	n.s.
Ergonomics	NOV <> EXP	<0.001*	EXP higher
Safety	NOV <> EXP	<0.001*	NOV higher
Fluid handling	NOV <> EXP	0.003*	EXP higher
Overall	NOV <> EXP	0.402	n.s.

* Significant difference between novices and experienced group ($p < 0.05$, Mann–Whitney *U*-test, two-sided, exact)

n.s., not significant

Learning curves

Figure 4 shows the learning curve via box plots of trial 1 to trial 5 for novices and experienced surgeons in the exploration and diagnosis exercises. The scores for the individual modules in the exploration exercise are shown in Fig. 5. As indicated in the "Construct validity" section above, the scores here are again similar for the visualization module, whereby the experienced group scored higher in the ergonomics module, while novices had significantly higher ratings for the safety module. Figure 6 depicts the learning curves of the individual modules for the diagnosis exercise, which was conducted directly after the exploration exercise.

The results of the training effect calculations are shown in Table 4. In the exploration exercise, the experienced group improved from both trial 1 and trial 2 to trial 5 for all modules except the safety module. Novices improved from trial 1 to trial 5 in all modules and in trial 2 to 5 for all modules except the visualization module.

Results were different in the more difficult diagnosis exercise, which requires fluid handling skills and presented a more complex anatomy with a larger pathology. Here, the experienced group did not improve significantly from trial 2 to trial 5 in any of the modules, while novices improved in the visualization module ($p = 0.002$) and in overall score ($p = 0.015$).

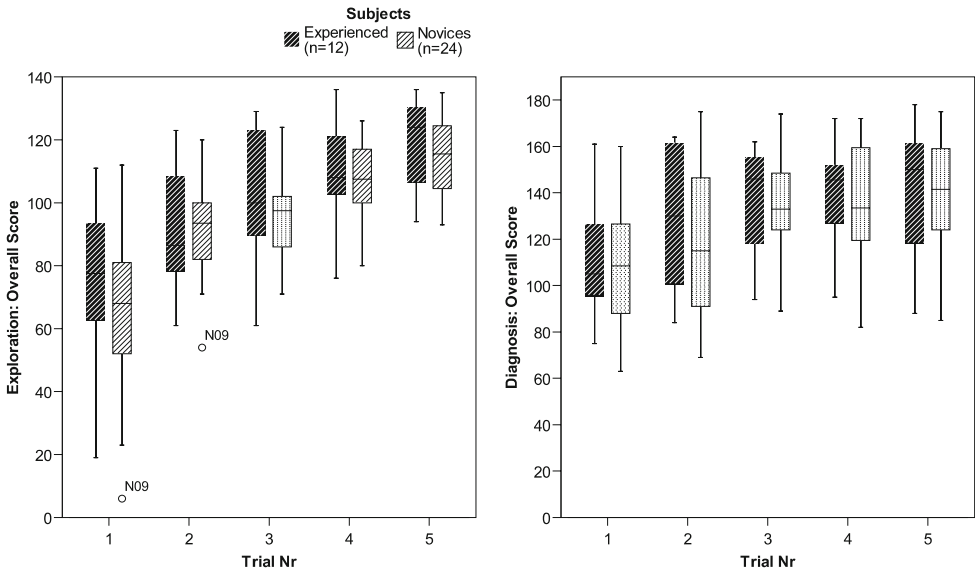

Fig. 4 Overall scores for the exploration exercise (*left*) and diagnosis exercise (*right*) for the experienced group (E01 to E12) and the novices (N01 to N24) from trial 1 to trial 5 as boxplots

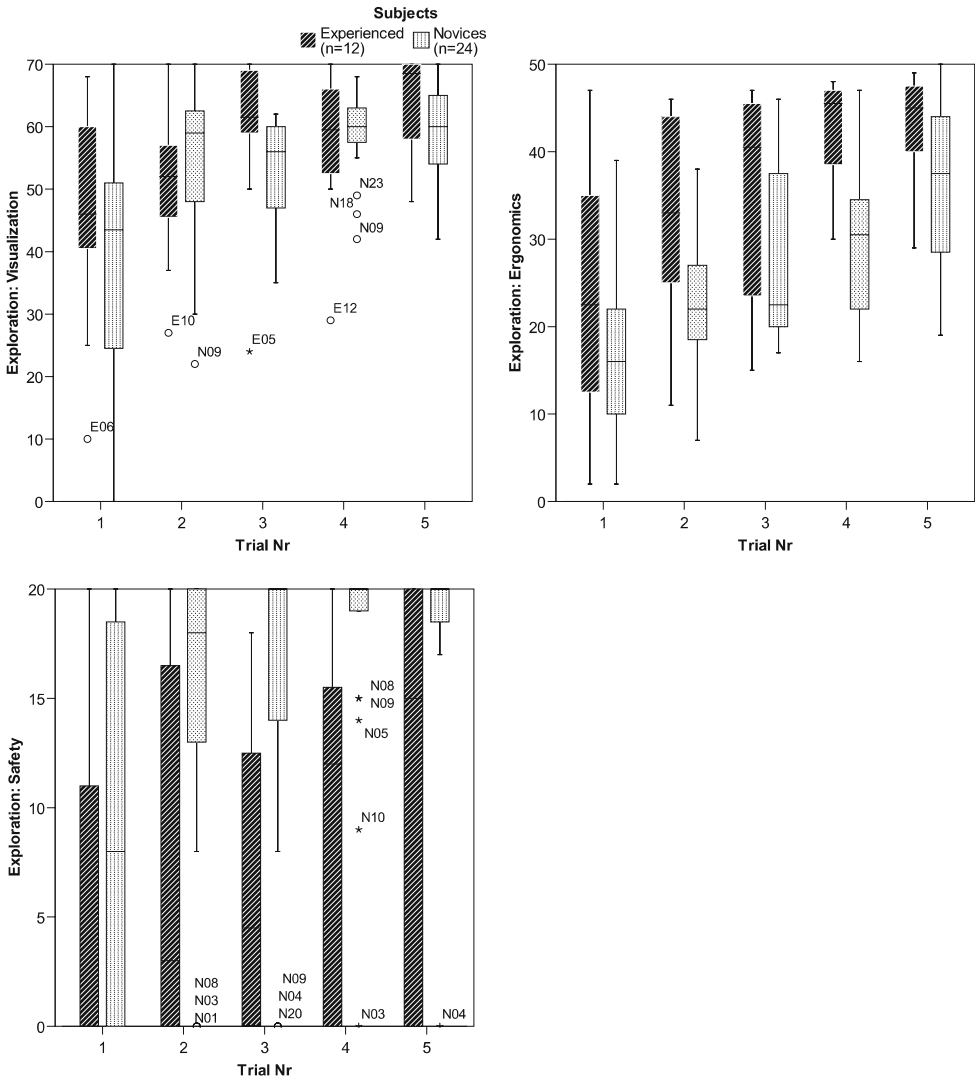

Fig. 5 Exploration exercise: experienced (E01 to E12) and novice scores (N01 to N24) for the visualization, ergonomics, and safety modules from trial 1 to trial 5 as boxplots

Repeated practice

In Fig. 7, the performance of the 23 novices who returned for repeated practice is displayed for both the exploration and diagnosis exercise. Trial 6 to 10 denote the second training session. In the exploration exercise, performance was significantly higher on trial 6 than trial 1 ($p < 0.001$), but dropped slightly from trial 5 to trial 6. From trial 6 it improved consistently until it reached a plateau with trial 9 and 10. For the diagnosis exercise, there was no drop between trial 5 and trial 6; however the total score increased significantly from trial 6 to trial 10 ($p = 0.012$).

We also investigated whether there were differences between the different scoring modules, e.g., if there was just a performance drop for the visualization module, but not for the safety module, between trial 5 and 6. However, we could not find any obvious and coherent relation.

Discussion

In this study, we introduced a new MMSS for the HystSim hysteroscopy training simulator, testing it on two different typical diagnostic procedures. While construct validity was shown for two of four scoring modules (ergonomics and fluid handling), the experienced group did not score significantly higher in the visualization module than novices, and the safety module showed a significant but inverse difference, with novices scoring higher than the experienced group. The surprising results in the safety module and also in the visualization module demand a more detailed analysis.

So far, the safety module has consisted of only one metric, namely the time the surgical tool was colliding with the uterine wall. It did not distinguish between critical and noncritical contact, i.e., the simulator did not take into account the penetration depth of the tool or the position or angle of the collision. Therefore, also lengthy yet noncritical

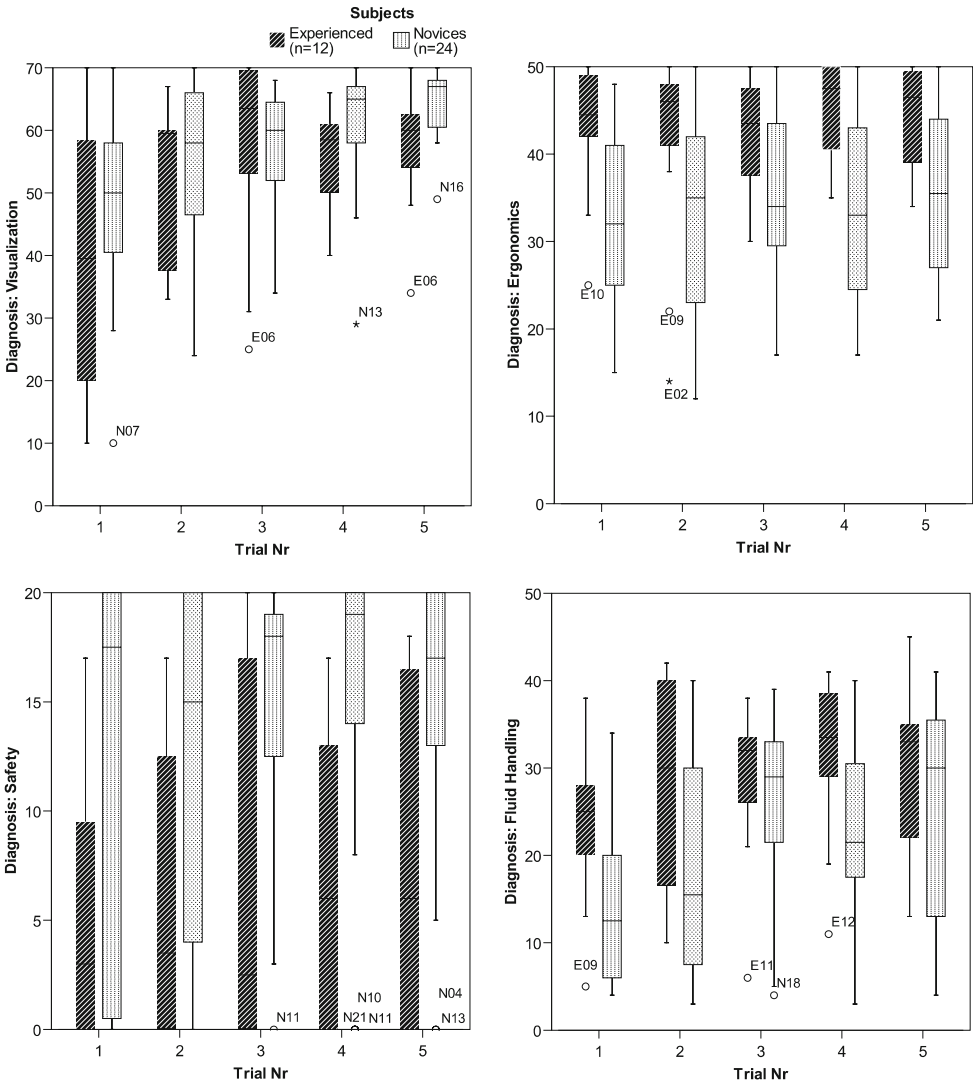

Fig. 6 Diagnosis exercise: experienced (E01 to E12) and novice scores (N01 to N24) for the visualization, ergonomics, safety, and fluid handling modules from trial 1 to trial 5 as boxplots

collisions, as often encountered during real interventions, lowered the safety module score substantially. We suppose that the experienced group performed surgery on the HystSim as during real hysteroscopy, thereby overlooking the clearly stated goal not to collide with the wall. Additionally, even if hysteroscopy is generally guided by vision, in this point the experienced gynecologists might have been misled by the missing haptic feedback. A post analysis of the overall scores without the safety module resulted in a significantly higher score for the experienced group than for the novices for both the exploration exercise ($p = 0.002$) and the diagnosis exercise ($p = 0.007$).

Nevertheless, we believe that careful handling and proceeding during hysteroscopy should be taught as part of a constitutive training since the potential to cause harm to patients changes dramatically when an operative element for resection under electricity is used.

Furthermore, according to self-declaration, none of our experienced users had ever been through a standardized curriculum or formal teaching for hysteroscopy when learning these procedures. This resulted in different approaches taken by the experienced surgeons for the two tasks. Even though the HystSim was able to handle all these different techniques while maintaining a realistic simulation, every trial was scored and graded according to the expert opinion as defined in the MMSS. Therefore, techniques clearly deviating from the assumed standard resulted in lower ratings. Particularly the visualization module and again the safety module were affected in this regard.

Therefore, it is recommended that these measurement modules should be further developed and refined before being used for providing feedback or evaluation of surgical performance.

Concerning learning curves, we found that all subjects improved significantly during the training on the simulator. However, some of the subjects reached a plateau with only ten repetitions. The effect of a performance plateau after a comparably small number of repetitions or training

Table 4 Statistical analysis of the training effect from trial 1 and trial 2 to trial 5 for both novices (NOV) and the experienced group (EXP)

Module	Comparison trials	*P*-value NOV	*P*-value EXP
Exploration exercise			
Visualization	1 versus 5	<0.001*	0.002*
Ergonomics	1 versus 5	<0.001*	0.002*
Safety	1 versus 5	<0.001*	n.s.
Overall	1 versus 5	<0.001*	<0.001*
Visualization	2 versus 5	n.s.	0.003*
Ergonomics	2 versus 5	<0.001*	0.017*
Safety	2 versus 5	0.022*	n.s.
Overall	2 versus 5	<0.001*	0.001*
Diagnosis exercise			
Visualization	1 versus 5	<0.001*	0.039*
Ergonomics	1 versus 5	n.s.	n.s.
Safety	1 versus 5	n.s.	n.s.
Fluid handling	1 versus 5	0.003*	n.s.
Overall	1 versus 5	<0.001*	0.024*
Visualization	2 versus 5	0.002*	n.s.
Ergonomics	2 versus 5	n.s.	n.s.
Safety	2 versus 5	n.s.	n.s.
Fluid handling	2 versus 5	n.s.	n.s.
Overall	2 versus 5	0.015*	n.s.

* Significant difference between trial 5 and trial 1 or 2 (trial 5 higher, $p < 0.05$, Mann–Whitney *U*-test, two-sided, exact)

n.s., not significant

sessions has been shown in studies for other surgical simulators, e.g., for laparoscopy [17, 34–36], where the plateau was reached mostly after only three to five repetitions on the simulator.

An explanation for the early plateau might be related to the setup of the experiment. At the time of the study, only the diagnostic hysteroscopy module was available in the simulator. While diagnostic hysteroscopy is considered as a procedure of low complexity which can easily be performed by novices under supervision, more complex interventions such as large myoma removal, endometrium resection or removal of large synechiae would be more challenging to novices and the experienced group alike, therefore probably prolonging the learning curve.

So far the assumed performance goals were defined by only two experts. When building a metric system for future procedures, the elaboration, choice, weighting, implementation, and configuration of the metrics should be established on a broader base with consensus on the level of a national or, even better, international taskforce.

Prior to a widely accepted integration of the HystSim training into the medical curriculum, the validation cascade will have to be completed with studies on predictive validity. The encouraging initial results by using HystSim for surgical skills training suggest that a curriculum for

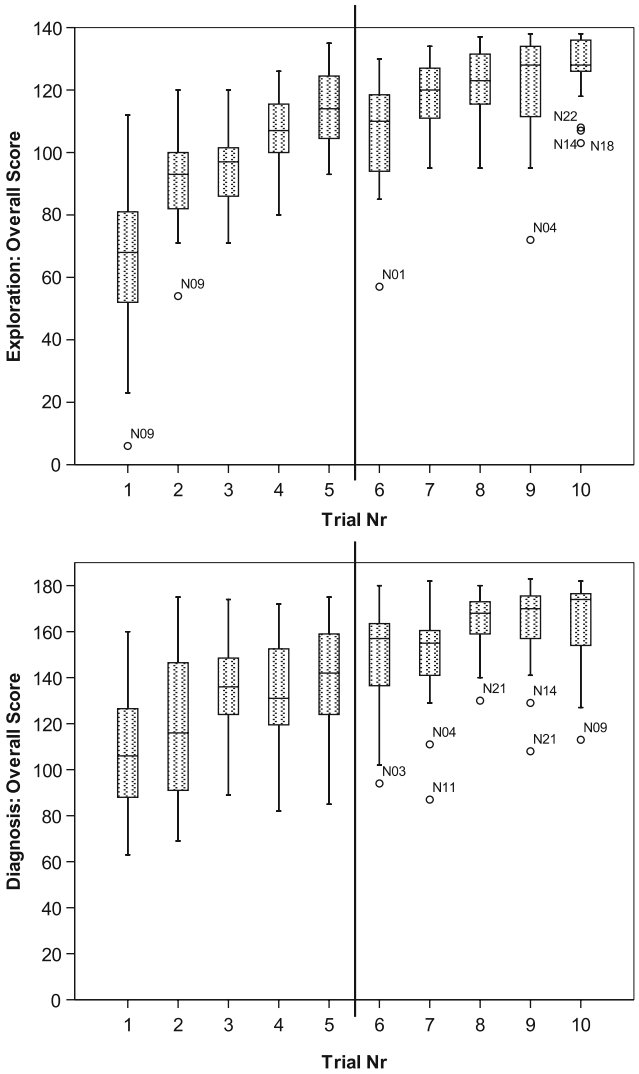

Fig. 7 Overall scores for the first training session (trials 1 to 5) and the second training session (trials 6 to 10) for the exploration exercise (*top*) and for the diagnosis exercise (*bottom*), novices only (N01 to N24)

hysteroscopy based on VR surgical training might be equally beneficial to both trainees and trainers, and ultimately, to patient safety.

Acknowledgment The authors would like to thank all developers of the hysteroscopy simulator project. This research has been supported by the NCCR Co-Me of the Swiss National Science Foundation. Funding: NCCR Co-Me (Computer Aided and Image Guided Medical Interventions) of the Swiss National Science Foundation (http://www.co-me.ch).

References

1. Aggarwal R, Ward J, Balasundaram I, Sains P, Athanasiou T, Darzi A (2007) Proving the effectiveness of virtual reality simulation for training in laparoscopic surgery. Ann Surg 246:771–779

2. Khalifa YM, Bogorad D, Gibson V, Peifer J, Nussbaum J (2006) Virtual reality in ophthalmology training. Surv Ophthalmol 51:259–273
3. Michelson JD (2006) Simulation in orthopaedic education: an overview of theory and practice. J Bone Joint Surg Am 88: 1405–1411
4. Panait L, Bell RL, Roberts KE, Duffy AJ (2008) Designing and validating a customized virtual reality-based laparoscopic skills curriculum. J Surg Educ 65:413–417
5. Park J, MacRae H, Musselman LJ, Rossos P, Hamstra SJ, Wolman S, Reznick RK (2007) Randomized controlled trial of virtual reality simulator training: transfer to live patients. Am J Surg 194:205–211
6. Wignall GR, Denstedt JD, Preminger GM, Cadeddu JA, Pearle MS, Sweet RM, McDougall EM (2008) Surgical simulation: a urological perspective. J Urol 179:1690–1699
7. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, Satava RM (2002) Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg 236:458–463
8. Schijven MP, Jakimowicz JJ, Broeders IA, Tseng LN (2005) The Eindhoven laparoscopic cholecystectomy training course—improving operating room performance using virtual reality training: results from the first E.A.E.S. accredited virtual reality trainings curriculum. Surg Endosc 19:1220–1226
9. Carter FJ, Schijven MP, Aggarwal R, Grantcharov T, Francis NK, Hanna GB, Jakimowicz JJ (2005) Consensus guidelines for validation of virtual reality surgical simulators. Work Group for Evaluation and Implementation of Simulators and Skills Training Programmes. Surg Endosc 19:1523–1532
10. Bajka M, Tuchschmid S, Streich M, Fink D, Székely G, Harders M (2008) Evaluation of a new virtual-reality training simulator for hysteroscopy. Surg Endosc 2008 Apr 24 [Epub ahead of print]
11. Harders M, Bachofen D, Bajka M, Grassi M, Heidelberger B, Sierra R, Spaelter U, Steinemann D, Teschner M, Tuchschmid S, Zatonyi J, Szekely G (2008) Virtual reality based simulation of hysteroscopic interventions. Presence-Teleop Virt Environ 17(5):441–462
12. Duffy AJ, Hogle NJ, McCarthy H, Lew JI, Egan A, Christos P, Fowler DL (2005) Construct validity for the LAPSIM laparoscopic surgical simulator. Surg Endosc 19:401–405
13. Eriksen JR, Grantcharov T (2005) Objective assessment of laparoscopic skills using a virtual reality stimulator. Surg Endosc 19:1216–1219
14. Gallagher AG, Richie K, McClure N, McGuigan J (2001) Objective psychomotor skills assessment of experienced, junior, and novice laparoscopists with virtual reality. World J Surg 25:1478–1483
15. McCloy R, Stone R (2001) Science, medicine, and the future. Virtual reality in surgery. BMJ 323:912–915
16. Satava RM (1993) Virtual reality surgical simulator. The first steps. Surg Endosc 7:203–205
17. Schijven M, Jakimowicz J (2003) Construct validity: experts and novices performing on the Xitact LS500 laparoscopy simulator. Surg Endosc 17:803–810
18. Sherman V, Feldman LS, Stanbridge D, Kazmi R, Fried GM (2005) Assessing the learning curve for the acquisition of laparoscopic skills on a virtual reality simulator. Surg Endosc 19:678–682
19. Taffinder N, Sutton C, Fishwick RJ, McManus IC, Darzi A (1998) Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: results from randomised controlled studies using the MIST VR laparoscopic simulator. Stud Health Technol Inform 50:124–130
20. Gallagher AG, Satava RM (2002) Virtual reality as a metric for the assessment of laparoscopic psychomotor skills Learning curves and reliability measures. Surg Endosc 16(12):1746–1752
21. Moorthy K, Munz Y, Dosis A, Bello F, Darzi A (2003) Motion analysis in the training and assessment of minimally invasive surgery. Minim Invasive Ther Allied Technol 12:137–142
22. Ritter EM, McClusky DA, Gallagher AG, Smith CD (2005) Real-time objective assessment of knot quality with a portable tensiometer is superior to execution time for assessment of laparoscopic knot-tying performance. Surg Innov 12:233–237
23. Rosen J, Hannaford B, Richards CG, Sinanan MN (2001) Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills. IEEE Trans Biomed Eng 48:579–591
24. Rosen J, Brown JD, Chang L, Sinanan MN, Hannaford B (2006) Generalized approach for modeling minimally invasive surgery as a stochastic process using a discrete markov model. IEEE Trans Biomed Eng 53:399–413
25. Van Dongen KW, Tournoij E, van der Zee DC, Schijven MP, Broeders IA (2007) Construct validity of the LapSim: can the LapSim virtual reality simulator distinguish between novices and experts? Surg Endosc 21:1413–1417
26. Mackay S, Datta V, Chang A, Shah J, Kneebone R, Darzi A (2003) Multiple Objective Measures of Skill (MOMS): a new approach to the assessment of technical ability in surgical trainees. Ann Surg 238:291–300
27. Heinrichs WL, Lukoff B, Youngblood P, Dev P, Shavelson R, Hasson HM, Satava RM, McDougall EM, Wetter PA (2007) Criterion-based training with surgical simulators: proficiency of experienced surgeons. JSLS 11:273–302
28. Cao CG, MacKenzie CL, Ibbotson JA, Turner LJ, Blair NP, Nagy AG (1999) Hierarchical decomposition of laparoscopic procedures. Stud Health Technol Inform 62:83–89
29. Tuchschmid S, Bajka M, Bachofen D, Székely G, Harders M (2007) Objective surgical performance assessment for virtual hysteroscopy. Stud Health Technol Inform 125:473–478
30. Gallagher AG, Ritter EM, Satava RM (2003) Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. Surg Endosc 17:1525–1529
31. Gallagher AG, Ritter EM, Champion H, Higgins G, Fried MP, Moses G, Smith CD, Satava RM (2005) Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. Ann Surg 241:364–372
32. Schijven MP, Jakimowicz J (2004) The learning curve on the Xitact LS 500 laparoscopy simulator: profiles of performance. Surg Endosc 18:121–127
33. Hassan I, Weyers P, Maschuw K, Dick B, Gerdes B, Rothmund M, Zielke A (2006) Negative stress-coping strategies among novices in surgery correlate with poor virtual laparoscopic performance. Br J Surg 93:1554–1559
34. Aggarwal R, Grantcharov TP, Eriksen JR, Blirup D, Kristiansen VB, Funch-Jensen P, Darzi A (2006) An evidence-based virtual reality training program for novice laparoscopic surgeons. Ann Surg 244:310–314
35. Chaudhry A, Sutton C, Wood J, Stone R, McCloy R (1999) Learning rate for laparoscopic surgical skills on MIST VR, a virtual reality simulator: quality of human-computer interface. Ann R Coll Surg Engl 81:281–286
36. Gor M, McCloy R, Stone R, Smith A (2003) Virtual reality laparoscopic simulator for assessment in gynaecology. BJOG 110:181–187